

# **Agricultural Health Study**



## **Resolved Reproductive Variables**

### **Data File Users Manual**

**Release REL0908.00**

Stanley E. Legum, Ph.D., Stuart Long, B. S., Editors

**December 2009**

## TABLE OF CONTENTS

<u>Chapter</u>		<u>Page</u>
1	INTRODUCTION .....	1
2	NEED FOR RESOLUTIONS.....	1
3	METHODS .....	2
4	LIST OF VARIABLES IN MENSTRUAL DATA SET.....	8
5	VARIABLE FORMATS .....	9

## 1. INTRODUCTION

The Agricultural Health Study (AHS) is a collaborative effort involving the National Cancer Institute (NCI), the National Institute of Environmental Health Sciences (NIEHS), and the U.S. Environmental Protection Agency (EPA). The goals are to investigate the effects of environmental, occupational, dietary, and genetic factors on the health of the agricultural population. This study is providing information that agricultural workers can use in making decisions about their health and the health of their families.

## 2. NEED FOR RESOLUTIONS

During Phase 1 of the AHS (between December 1993 and December 1997) 52,394 private applicators and 32,391 of their spouses enrolled AHS. Of the private applicators, 1,359 were female, while 32,172 of the 32,391 spouses were female. The females who enrolled in the study were further asked to complete the Female and Family Health questionnaire (FFH). This questionnaire included extensive questions about menopause and menstrual issues.

During Phase 2 of the AHS (from July 1999 through August 2003 for private applicators and their spouse and from October 2003 through October 2005 for commercial applicators) Computer Assisted Telephone Interview (CATI) with members of the cohort enrolled during Phase 1. Of the 1,359 female private applicators, 551 completed the FFH, while 122 of the 204 female commercial applicators completed the FFH, and 19,623 of the 32,172 female spouses completed the FFH. Of the 808 female private applicators without an FFH, 502 completed the Phase 2 interview. Of the 12,549 female spouses without an FFH, 8,717 completed the Phase 2 interview. Of the 82 female commercial applicators without an FFH, 36 completed the Phase 2 interview.<sup>1</sup>

This means that we have 9,255 females from Phase 1 without a completed FFH who have Phase 2 CATI data. The Phase 2 CATI included a health section asking about menopause and menstrual issues.

---

<sup>1</sup> Resolved reproductive variables have not yet been constructed for commercial applicators. The file contains resolved variables only for private applicators and spouses of private applicators.

The answers provided in the Phase 2 CATI can be used to fill in missing data from the Phase 1 period, and consequently can be used in future analyses. In addition, responses to menopause and menstrual issues from Phase 1 can be used to fill in missing data in the Phase 2 CATI.

### 3. METHODS

Several AHS data sets were used to resolve the data. The Demographic File (release AHSREL0905.00) was used to retrieve genders, participant applicator or spouse status, birth dates and questionnaires completed. These data were subset to contain only female respondents (private applicators, commercial applicators and spouses combined). Phase 1 menopause and menstrual data was retrieved from the FFH File. Phase 2 menopause and menstrual data were retrieved from the Phase 2 Spouse File, the Phase 2 Private Applicator File and the Phase 2 Commercial Applicator File.

After combining the demographic, Phase 1 and Phase 2 data, the following actions were performed to fill in missing data and to resolve discrepancies in the data where possible:

#### **Age at Last Menstrual Period (LMP).**

The following steps were used to create the variable LMPAGE (age at last menstrual period):

1. Phase 2 unknown (..9), refusal (..8) and erroneous data were treated as missing values.
2. Depending on the Phase 2 female health questionnaire version administered<sup>2</sup>, women provided either an age at LMP or a date of LMP. Since these two items were mutually exclusive, all of the dates of LMP were converted to an age at LMP for the Resolved Reproductive Variable data set.
3. The Phase 2 LMP fix data was then incorporated into the Phase 2 age at LMP. To do this, if there was an age recorded in the S2\_LMP\_FIX variable that was different from the age recorded in the S2\_LMPAGE variable, then the S2\_LMPAGE variable was set to the value recorded in the S2\_LMP\_FIX .

---

<sup>2</sup> See the *Phase II Data File Users Manual* for details.

4. Certain improbable Phase 1 and Phase 2 responses were resolved by setting their values to missing. If a woman gave her age at LMP as greater than her current or gave her age at LMP as greater than 70, the age at LMP was set to missing. There were less than 100 changes of this type
5. Phase 1 and Phase 2 age at LMP were combined to create the age at LMP variable (LMPAGE). If the Phase 2 LMP existed, the combined age variable was set to the Phase 2 value. Otherwise, LMPAGE was set to the Phase 1 value.

### **Ever Gone Through Menopause.**

Conflicts between Phase 1 menopause responses and Phase 2 menopause or LMP ages were resolved, if possible. The results were used to create the variables P1\_HRT\_EVER and P2\_HRT\_EVER, for Phases 1 and 2, respectively.

1. If the Phase 1 menopause response was YES and the Phase 2 menopause response was missing, then we changed the Phase 2 menopause response to YES.
2. If the Phase 2 menopause response was NO and the Phase 1 menopause was missing or unsure, then we changed the Phase 1 menopause response to NO.
3. If the Phase 1 menopause response was YES and the Phase 2 menopause response was NO, then the record was reviewed to determine if a resolution of the conflict could be made. Many of the conflicts were due to Phase 1 surgery types incorrectly being identified as surgeries that cause surgical menopause. In these situations, the Phase 1 menopause response was changed from YES to NO.

Menopause status was then categorized into 10 status levels for both Phase 1 (variable P1\_MENOPAUSE\_STATUS) and Phase 2 (variable P2\_MENOPAUSE\_STATUS):

- |   |                                       |
|---|---------------------------------------|
| 0 | No Menopause                          |
| 1 | Natural Menopause                     |
| 2 | Hysterectomy and both ovaries removed |
| 3 | Both ovaries (only) removed           |
| 4 | Other type of Menopause               |

- 5 Uterus and/or one ovary removed where age > 55
- 6 Surgery of unknown type where age > 55
- 7 Uterus and/or one ovary removed where age <= 55
- 8 Surgery of unknown type where age <= 55
- 9 Unknown

These 10 levels were collapsed to create a 5 level resolved menopause variables for each study phase (MENOPAUSE\_P1 and MENOPAUSE\_P2). The levels are:

- 0 No Menopause
- 1 Natural Menopause
- 2 Menopause, surgical or other
- 3 Menopause, unknown where age >55
- 4 Unknown

#### **Use of Birth Control Pills (BCP's).**

In Phase 1 women were asked the age at which they first took birth control pills (variable WAGEPILL) and the number of years that they took birth control pills (variable WYRSPILL). The age question was written so as to elicit an exact age. The number of years question (question 5) allowed responses of none or one of 6 age ranges.

**5. Not counting any time when you stopped, for how many years altogether did you take or use birth control pills?** *(Mark appropriate answer)*

- None
- Less than 1 year
- 1–3 years
- 4–5 years
- 6–10 years
- 11–15 years
- 16 or more years

The following rules were applied to create the combined birth control use variables using information from both Phase 1 and Phase 2:

1. In Phase 1, six women provided an age started using birth control pills that was older than their age at enrollment.. All six of these women provided a response to the question of how many

years total have you used BCP's. To obtain a corrected age of first use, we subtracted the midpoint of the category she marked in question 5, from her age of enrollment to get an estimated age of first use.

2. A skip pattern fix was applied for Phase 1. If a woman responded that she had used BCPs for any number of years, while at the same time she did not respond YES to ever used BCP (WBCPILLS), then the value of WBCPILLS was treated as if she had responded YES.
3. In Phase 1, total years used BCP's (WYRSPILL) was asked in categories. If the midpoint of the category added to the age first used was greater than the age at enrollment, then the total years used was set to the age at enrollment minus the age first used (WAGEPILL).
4. For the Phase 2 Women's Health Module C, if the woman indicated that she used BCP since enrollment, then the total years used BCP was set to the sum of the Phase 1 BCP years and the number of years she used BCP since enrollment. If there were no total years used BCP in Phase 2 then the total years used BCP was set to the total years she used BCP in Phase 1.
5. A Phase 2 ever used BCP variable (P2\_BCP\_EVER) was created using the following rules:
  - a. If the woman ever said that she used BCP (Phase 1 or Phase 2 Health Module A or C) then this variable was set to YES (1).
  - b. Otherwise, this variable was set to NO (0) if:
    - i. She said that she had not used BCP in Phase 2 Health Module C; or-
    - ii. She indicated in the Phase 1 FFH that she had not used BCP AND she indicated in Phase 2 Health Module A that she had not used BCP
  - c. Otherwise, P2\_BCP\_EVER was set to missing.

## Hormone Replacement Therapy (HRT).

In Phase 1 women were asked the age at which they first took hormones/estrogen (variable WAGEHORM) and the number of years that they took hormones/estrogen (variable WYRSHORM). The age question was written so as to elicit an exact age. The number of years question (question 12) allowed responses of none or one of 6 age ranges.

**12. How many years altogether did you take hormones/estrogen for symptoms or conditions associated with menopause?** *(Do not count any time when you stopped.)* [wyrshorm]

- None
- Less than 1 year
- 1–3 years
- 4–5 years
- 6–10 years
- 11–15 years
- 16 or more years

1. In Phase 1, 23 women provided an age started using HRT that was older than their age at enrollment. All 23 of these women provided a response to the question of how many years total they had used HRT. To obtain a correct age of first use, we subtracted the midpoint of the category she marked in question 12, from the age of enrollment to get an estimated age of first use.
2. A skip pattern fix was applied for Phase 1. If a woman responded that she had used HRT for any number of years, while at the same time she did not respond YES to ever used HRT (WHORMONE), then we treated the value of WHORMONE as if she had responded YES.
3. In Phase 1, total years used HRT (WYRSHORM) was asked in categories. If the midpoint of the category added to the age first used was greater than the age at enrollment, then the total years used was set to the age at enrollment minus the age first used.
4. For Phase 2 Women's Health Module C, if the woman indicated that she used HRT since enrollment, then the total years used HRT was set to the sum of the Phase 1 HRT years and the number of years she used HRT since enrollment. If there were no total years used HRT in Phase 2 then the total years used HRT was set to the total years used HRT in Phase 1.
5. A Phase 2 ever used HRT variable (P2\_HRT\_EVER) was created using the following rules:
  - a. If the woman said that she used HRT (Phase 1 or Phase 2 Health Module A or C), then this variable was set to YES (1).

- b. Otherwise P2\_HRT\_EVER was set to NO (0) if
  - i. She said that she had not used HRT in Phase 2 Health Module C; or
  - ii. She indicated in the Phase 1 FFH that she had not used HRT AND she indicated in Phase 2 Health Module B that she had not used HRT
- c. Otherwise P2\_HRT\_EVER was set to missing.

### **Estrogen/Progesterone (EP).**

The variables P1\_PROGEST\_EVER and P2\_PROGEST\_EVER were created using the following rules:

1. Set P1\_PROGEST\_EVER = to the value of the Phase 1 variable WPROGEST.
2. In Phase 2, if the woman did not use HRT, then consider the number of EP years to be 0 and set P2\_PROGEST\_EVER = 0 (NO).
3. When the Phase 1 EP value is YES and the Phase 2 EP value is missing, then set P2\_PROGEST\_EVER = 1 (YES).
4. If the woman responded in Phase 2 that she had never used EP, but the Phase 1 response was positive for EP then use the Phase 2 response (i.e., P2\_PROGEST\_EVER = 0) and set P1\_PROGEST\_EVER = 0 (NO).
5. In both phases, if the woman had a missing response for HRT but answered NO for EP, then, following the skip pattern in the questionnaire, the EP usage value was treated as missing for that phase (P1\_PROGEST\_EVER = . (missing) and P2\_PROGEST\_EVER = . (missing)).
6. If the woman was missing EP use at Phase 1 but had used EP at the time of the Phase 2 interview for a number of years that predates the Phase 1 interview, then treat Phase 1 EP usage YES and set P1\_PROGEST\_EVER = 1.
7. If the Phase 2 HRT response was NO, then, following the skip pattern, set P2\_PROGEST\_EVER = 0 (NO)

#### 4. LIST OF VARIABLES IN THE MENSTRUAL DATA SET

Variable Name	Label
PARTID	Participant Identifier
AP_SPOUSE	Type of AHS Participant
LMPAGE	Phase 1/Phase 2: Age of Last Menstrual Period (LMP) resolved for Phase 1 LMP age, Phase 2 LMP age and Phase 2 Date of Last Menstrual Period
P2_MENOPAUSE_STATUS	Phase 1/Phase 2: Phase 2 Ten category menopause status resolved with Phase 1 and Phase 2 data
P1_MENOPAUSE_STATUS	Phase 1/Phase 2: Phase 1 Ten category menopause status resolved with Phase 1 and Phase 2 data
MENOPAUSE_P1	Phase 1/Phase 2: Phase 1 Resolved Menopause variable (5 level)
MENOPAUSE_P2	Phase 1/Phase 2: Phase 2 Resolved Menopause variable (5 level)
P1_BCP_AGE_FIRST_USE	Phase 1 age first used Birth Control Pill, resolved with Phase 1 data
P1_BCP_EVER	Phase 1/Phase 2: Phase 1 Birth Control Pill use resolved with Phase 1 and Phase 2 data
P1_BCP_YEARS	Phase 1 years used Birth Control Pills, resolved with Phase 1 data
P2_BCP_EVER	Phase 1/Phase 2: Phase 2 Birth Control Pill use resolved with Phase 1 and Phase 2 data
P2_BCP_YEARS	Phase 1/Phase 2: Phase 2 years used Birth Control Pills resolved with Phase 1 and Phase 2 data
P1_HRT_AGE_FIRST_USE	Phase 1 age first used Hormone Replacement Therapy, resolved with Phase 1 data
P1_HRT_EVER	Phase 1/Phase 2: Phase 1 Hormone Replacement Therapy resolved with Phase 1 and Phase 2 data
P1_HRT_YEARS	Phase 1 years used Hormone Replacement Therapy, resolved with Phase 1 data
P2_HRT_EVER	Phase 1/Phase 2: Phase 2 Hormone Replacement Therapy resolved with Phase 1 and Phase 2 data
P2_HRT_YEARS	Phase 1/Phase 2: Phase 2 years used Hormone Replacement Therapy resolved with Phase 1 and Phase 2 data
P1_PROGEST_EVER	Phase 1/Phase 2: Phase 1 Progesterone with Estrogen use resolved with Phase 1 and Phase 2 data
P2_PROGEST_EVER	Phase 1/Phase 2: Phase 2 Progesterone with Estrogen use resolved with Phase 1 and Phase 2 data

## 5. VARIABLE FORMATS

```
PROC FORMAT ;
  VALUE $as
    "C" = "Commercial Applicator"
    "P" = "Private Applicator"
    "S" = "Spouse";
  VALUE _meno
    0 = "No Menopause"
    1 = "Yes-Natural"
    2 = "Yes-Surgical/Other"
    3 = "Yes-Unknown age>55"
    4 = "Unknown";
  VALUE _Status
    0 = "No Menopause"
    1 = "Yes, natural"
    2 = "Yes, Hysterectomy/both ovaries"
    3 = "Yes, both ovaries (only)removed"
    4 = "Yes, other type"
    5 = "Yes, uterus and/or one ovary, age>55"
    6 = "Yes, surgery unknown type, age>55"
    7 = "Yes, uterus and/or one ovary, age<=55"
    8 = "Yes, surgery unknown type, age<=55"
    9 = "unknown";
RUN;
```